Title
Chemoinformatic-Guided Engineering of Polyketide Synthases.

Journal
Journal of the American Chemical Society, 142(22)

ISSN
0002-7863

Authors
Zargar, Amin
Lal, Ravi
Valencia, Luis
et al.

Publication Date
2020-06-01

DOI
10.1021/jacs.0c02549

Peer reviewed

Chemoinformatic-Guided Engineering of Polyketide Synthases

Amin Zargar, Ravi Lal, Luis Valencia, Jessica Wang, Tyler William H. Backman, Pablo Cruz-Morales, Ankita Kothari, Miranda Werts, Andrew R. Wong, Constance B. Bailey, Arthur Loubat, Yuzhong Liu, Yan Chen, Samantha Chang, Veronica T. Benites, Amanda C. Hernańdez, Jesus F. Barajas, Mitchell G. Thompson, Carolina Barcelos, Rasha Anayah, Hector Garcia Martin, Aindrila Mukhopadhyay, Christopher J. Petzold, Edward E. K. Baidoo, Leonard Katz, and Jay D. Keasling*

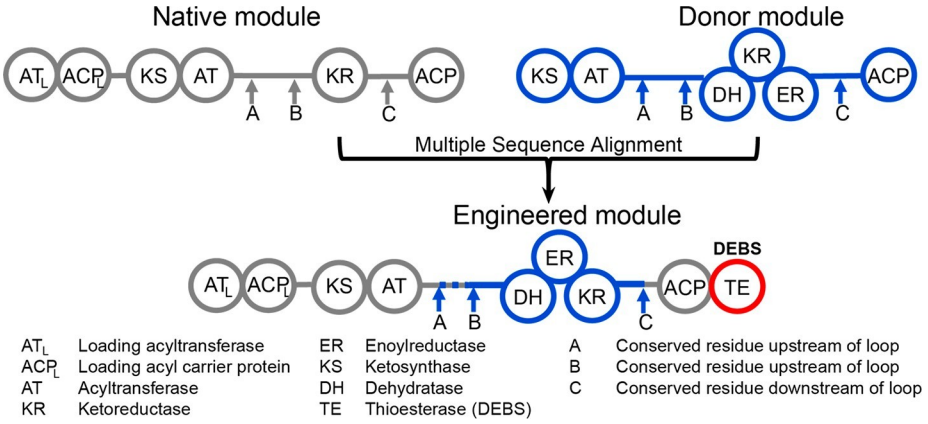
ABSTRACT: Polyketide synthase (PKS) engineering is an attractive method to generate new molecules such as commodity, fine and specialty chemicals. A significant challenge is re-engineering a partially reductive PKS module to produce a saturated β-carbon through a reductive loop (RL) exchange. In this work, we sought to establish that chemoinformatics, a field traditionally used in drug discovery, offers a viable strategy for RL exchanges. We first introduced a set of donor RLs of diverse genetic origin and chemical substrates into the first extension module of the lipomycin PKS (LipPKS1). Product titers of these engineered unimodular PKSs correlated with chemical structure similarity between the substrate of the donor RLs and recipient LipPKS1, reaching a titer of 165 mg/L of short-chain fatty acids produced by the host *Streptomyces albus* J1074. Expanding this method to larger intermediates that require bimodular communication, we introduced RLs of divergent chemosimilarity into LipPKS2 and determined triketide lactone production. Collectively, we observed a statistically significant correlation between atom pair chemosimilarity and production, establishing a new chemoinformatic method that may aid in the engineering of PKSs to produce desired, unnatural products.

Rational reprogramming of polyketide synthase (PKS) enzymes for the biosynthesis of new polyketides has been a major research thrust over the past three decades.[1−3] PKSs load a malonyl-CoA analog onto the acyl carrier protein (ACP) using the acyltransferase (AT) domain and extend the growing chain from the ketosynthase (KS) domain through a decarboxylative Claisen condensation reaction. After chain extension, the β-carbonyl reduction state is determined by the module's reductive domains, namely the ketoreductase (KR), dehydratase (DH), and enoylreductase (ER), which generate the β-hydroxyl, α,β-alkene, or saturated β-carbons, respectively, when progressively combined. Unlike fatty acid synthases, which faithfully produce saturated fatty acids, PKSs have this variability in β-carbonyl reduction. Consequently, multiple studies have reported PKS module engineering for various β-carbon oxidation states.[4−8] However, design strategies for introduction of reductive loop (RL) exchanges (i.e., KR-DH-ER domains) into partially reductive modules remain elusive. In this work, we compare bioinformatic and chemoinformatic approaches to guide RL exchanges and develop a new method for RL exchanges based on the chemical similarity of the RL substrate. Chemoinformatics, an interdisciplinary field blending computa- tional chemistry, molecular modeling and statistics to analyze structure−activity relationships, was first established for drug discovery.[9] Recently, we suggested that a chemoinformatic approach to PKS engineering could be valuable, particularly in RL exchanges where the KR and DH domains are substrate- dependent:[1] acyl chain length has critically affected dehydration in stand-alone DH[10−12] and full PKS module studies.[7,13]

Chemoinformatic methods such as atom pair (AP) similarity, which characterizes APs (e.g., length of bond path, number of π electrons), and maximum common substructure (MCS) similarity, which identifies the largest common substructure between two molecules,[14] could beneficially describe substrate profiles. While divergent in chemical characterization, both similarity methods translate to a Tanimoto coefficient with a range of 0 (least similar) to 1 (most similar).[14] We hypothesized that chemosimilarity between the substrates of donor and acceptor modules in RL exchanges may correlate with production levels, thereby leading to engineered modules that better control the reductive state of the β carbon.

Bioinformatic studies of PKS evolution have guided engineer- ing efforts in closely related biosynthetic gene clusters (BGCs).[15,16] We therefore undertook a phylogenetic analysis of the reductive domain common to all RLs, the KR. The KR not only reduces the β-keto group to a β-hydroxyl but also sets the stereochemistry of the β-group and, if a branched extender is used, sets the α-carbon stereochemistry resulting in subtypes A1, A2, B1, and B2 (Figure S1A). We generated a phylogenetic tree from all manually curated KRs and KSs in ClusterCAD, an

Scheme 1. Experimental Design of RL Swaps[a]



[a]Conserved residues are identified through multiple sequence alignment surrounding the reductive domains (A, B, and C). Donor RLs are inserted into the native lipomycin module 1, and the attached DEBS thioesterase hydrolyzes the product.

online database and toolkit for Type I PKSs, totaling 72 BGCs and 1077 modules.[17] As in previous investigations,[18,19] the KR domains clustered by subtype (Figures S1B and S2). In contrast, the RL type (e.g., KR, KR-DH, and KR-DH-ER) did not phylogenetically cluster with its upstream or downstream KS domain (Figures S3 and S4).[18] This suggests a link between KR evolution and product specificity, analogous to the evolution of KS domains of cis-AT[18] and trans-AT PKS modules[20,21] toward substrate specificity. As KRs from KR-DH-ER modules evolved distinctly from KR-only modules, we hypothesized that neither KR sequence identity nor phylogenetic distance, a pairwise comparison of phylogenetic tree members, between the donor loops and acceptor module were likely to correlate with RL exchange production levels.

To evaluate the importance of chemical similarity and phylogenetic distance in RL exchanges, we swapped diverse, full RLs into the first module of the lipomycin PKS (LipPKS1) using conserved residues as exchange sites (Scheme 1).[7] In our previous work, we introduced a heterologous thioesterase from 6-deoxyerythronolide B synthase (DEBS) into the C-terminus of LipPKS1; the resulting truncated PKS produced a β-hydroxy acid.[22] In this work, we selected N-terminal junctions (labeled A and B) located immediately after the post-AT linker, which is important for KS-AT domain architecture,[23] and the C-terminal junction (labeled C) directly before the ACP domain (see Table S1 for sequences) based on previous work with the first module of borreledin.[7]

We identified four donor RLs (IdmO, indanomycin, *S. antibioticus*; SpnB, spinosyn, *S. spinosa*; AurB, aureothin, *S. aureofaciens*; NanA2, nanchangamycin, *S. nanchangensis*; final products in Figure S5) to swap into LipPKS1. A pairwise comparison of phylogenetic distance and amino acid sequence identity determined that IdmO, AurB, and SpnB have the highest KR similarities to LipPKS1 (Figure 1A). A similar trend holds in the analysis of these donor modules upstream and downstream KS domains (Figure S6). In contrast, the NanA2 substrate has the highest chemical similarity based on AP and MCS similarity to LipPKS1, followed by SpnB (Figure 1B). With the introduction of RL swaps, the chimeric enzymes should produce 2,4-dimethylpentanoic acid. As *in vitro* PKS studies have shown divergence from *in vivo* results[24,25] due to underestimation of factors including limiting substrate, crowd- ing, and solubility,[26] we cloned eight chimeric modules and a control expressing red fluorescent protein (RFP), into an *E.*

coli−*Streptomyces albus* shuttle vector and conjugated into *S. albus* J1074 (Table S1).[27] Following ten-day production runs in a rich medium in biological triplicate, cultures of *S. albus* harboring each of the constructs were harvested and analyzed for product (Supporting Information).

Consistent with our hypothesis, we found a perfect correlation between titers of the desired product and the AP/MCS chemosimilarities between donor and LipPKS1 module substrates (R_s = 1.00 and p = 0.00) (Figure 1C). On the other hand, no significant correlation between product titer and phylogenetic distance or sequence similarity of the KR domain (R_s = 0.04, p = 0.60) was found. The lack of phylogenetic correlation was not surprising based on our bioinformatics analysis since the lipomycin KR is an A2-type, evolving separately from KRs with full RLs. This trend held in both junctions, though junction B chimeras generally resulted in higher product titers, consistent with a previous study of RL exchanges as the extra residues in junction A are distal to the ACP docking interface and active site.[7] Substituting the donor loop most chemically similar to LipPKS1, NanA2, resulted in the highest titers of desired product, 2,4-dimethylpentanoic acid, reaching 165 mg/L (Supporting Information). Low titers of the intermediate 2,4-dimethyl-3-hydroxypentanoic acid were pro- duced, which we hypothesize is due to a comparatively lower rate of turnover at the energetically intensive DH domain,[28] allowing for premature cleavage of the stalled product by non- enzymatic or TE-mediated hydrolysis. Like our previous study of *in vitro* production of adipic acid, we did not detect alkene or keto acid stalled products;[7] non-functional KRs produce short- chain β-keto acids that spontaneously decarboxylate to form ketones, which was also not observed, and ERs rapidly reduce *trans* double bonds.[28]

Based on these results, we took a chemoinformatic approach to further test our hypothesis that chemosimilarity of RL substrates is critical to PKS engineering. Using the Cluster- CAD[17] database, we identified donor RLs from laidlomycin and monensin that use a KR substrate (identical to the NanA2 KR substrate) with the highest chemically similarity to LipPKS1 (Figure 2A). As junction B resulted in superior levels of production, the RLs of LaidS2 and MonA2 were cloned into junction B of lipomycin. Like NanA2, LaidS2 loops produced high titers of desired product, while MonA2 performed similarly to SpnB and AurB (Figure 2B). As protein levels may influence product titers, we determined the quantitative levels of all

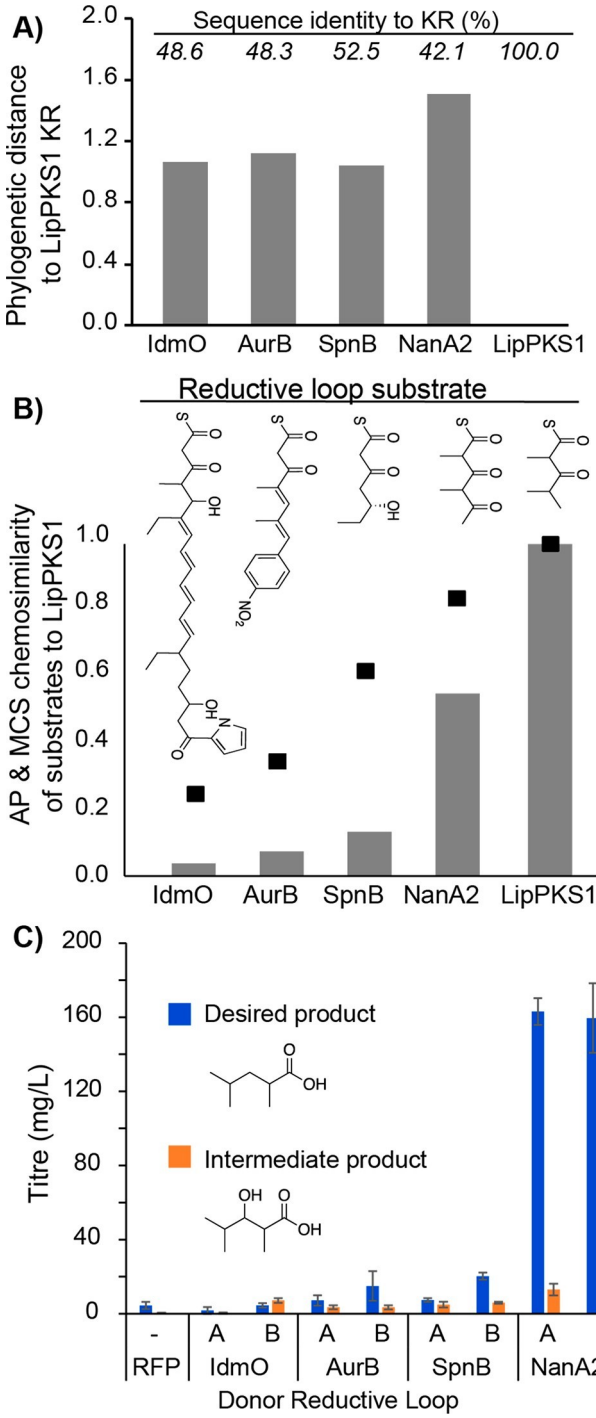
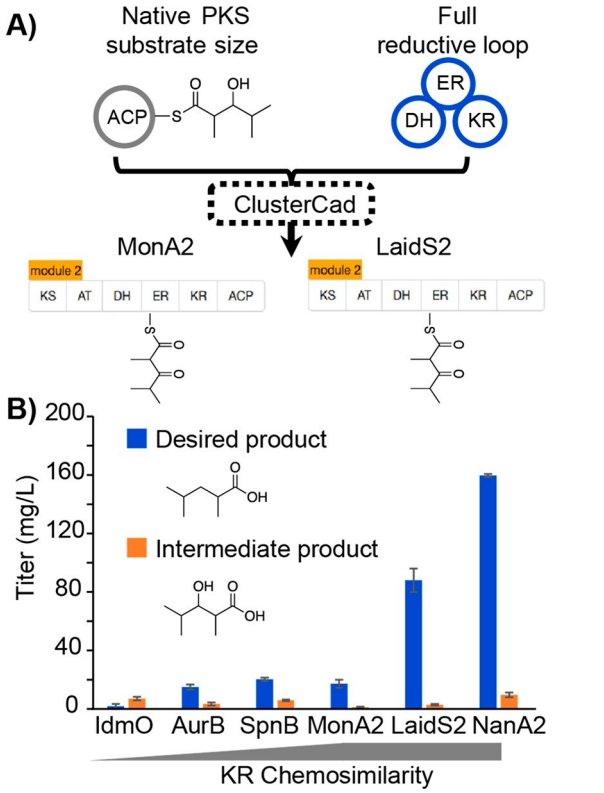


Figure 1. Phylogenetic and chemical similarity effects on reductive loop exchanges. (A) Phylogenetic distance of the native LipPKS1 KR domain to each donor KR. The value above each bar denotes KR sequence identity comparison. (B) AP (bar) and MCS (dots) chemical similarity between the native LipPKS1 KR domain and each donor KR. Chemical structures display native KR substrate in each module. (C) Polyketide production of engineered PKSs at both junctions A and B in biological triplicate (error bars denote standard deviation).

LipPKS1 constructs using targeted proteomics at the conclusion of the production run and observed no correlation between PKS protein levels and product titers (R_s = −0.15 and p =



0.77) (Figure S7). Reduced protein levels in the MonA2 swap could partially explain the lower levels of production in the MonA2 swap compared to LaidS2 and NanA2. However, targeted

Figure 2. A chemoinformatic approach to reductive loop exchanges.
(A) ClusterCad search revealed the closest substrates to LipPKS1 containing full RLs. (B) Production levels of junction B. RL exchanges are ordered from highest KR substrate similarity with LipPKS1 (MonA2, LaidS2, and NanA2) to progressively less similarity (IdmO, AurB, and SpnB) in biological triplicate (error bars denote standard deviation).

proteomics of three peptide peaks across the PKS does not eliminate the possibility of proteolytic degradation or variability in protein quality. AP Tanimoto and MCS chemosimilarity had equivalent Spearman rank correlation to product titers ($R_s = 0.82$, $p = 0.045$).

To better demonstrate the utility of this approach, we further evaluated RL exchanges where AP and MCS chemosimilarity diverge and tested this method in modules located at the center of assembly lines, thus requiring docking domain interactions and larger substrates. We therefore performed RL swaps on the second module of lipomycin, LipPKS2 (Figure 3A), to generate triketide lactones. Donor loops from SpnB and NanA2 were selected, as NanA2 has higher AP chemosimilarity while SpnB has higher MCS chemosimilarity (Figure 3B). As in our single-module swaps, KR phylogenetic similarity and sequence identity did not correlate with product titers. We found higher correlation with AP chemosimilarity due to higher product levels with NanA2 (Figure 3C,D). Proteomics on each PKS of these bimodular systems was not performed to rule out the effect of variable protein levels. AP chemosimilarity more heavily weights substructures, so NanA2 and LipPKS2 have higher similarity levels because both select methylmalonyl-CoA in the first two modules. In contrast, MCS chemosimilarity simply considers the largest common substructure, which ignores the influence of commonality at the growing chain by methyl groups. While extension of this phenomenon to account for variances in chemical similarity metrics (e.g., AP, MCS) requires further study, we hypothesize that chemosimilarity metrics that

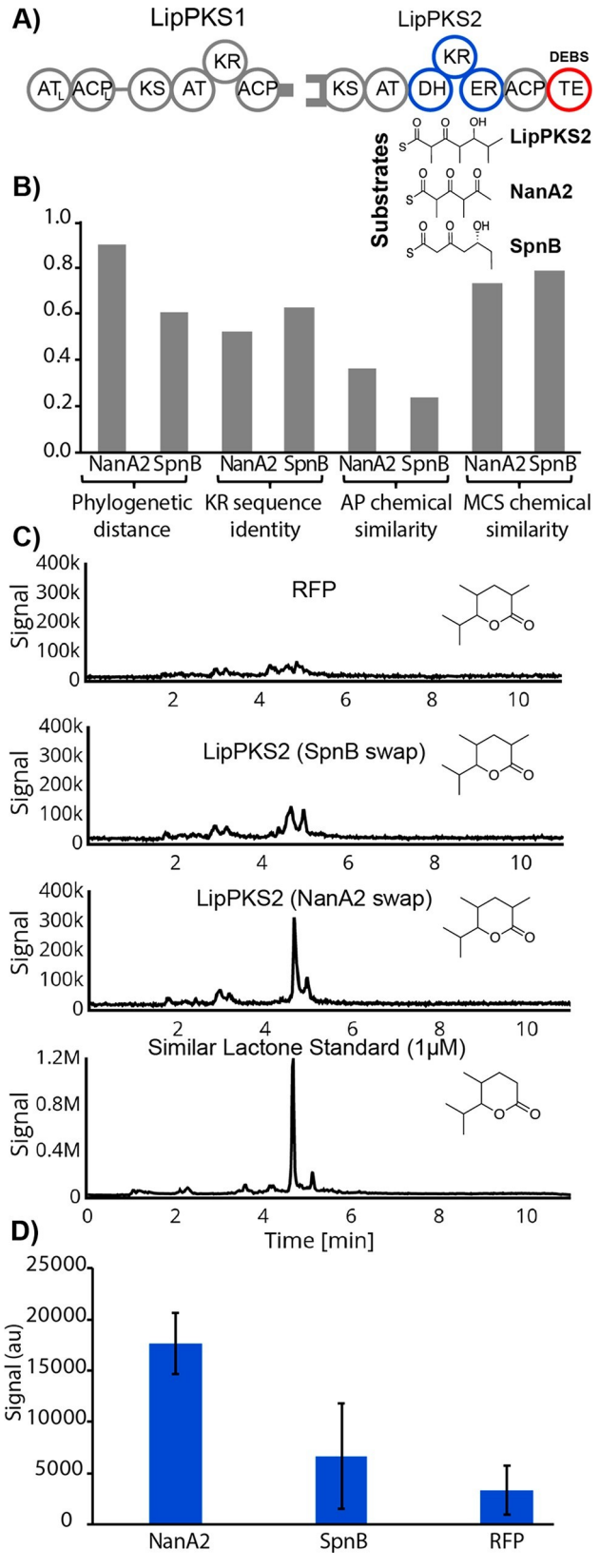


Figure 3. Bimodular reductive loop exchange. (A) Schematic of RL exchanges in LipPKS2 with substrates. (B) Phylogenetic distance, KR sequence identity, AP, and MCS similarity between RL donors and LipPKS2. (C) Chromatograms of RFP, LipPKS2 with donor loops SpnB and NanA2, and a structurally similar standard spiked into RFP cultures. (D) Production levels of desired lactone in biological triplicate (error bars denote standard deviation).

best match PKS enzymatic processing may prove most successful. Overall, in our RL exchanges in both LipPKS1 and LipPKS2 we determined a Spearman correlation between AP Tanimoto chemosimilarity and product titer to have an R_s = 0.88 and p = 0.004 (Supporting Information).

Based on previous literature regarding the importance of substrate size in reductive domains, in this study we hypothesized that the field of chemoinformatics, traditionally used in drug discovery, could be applied to PKS engineering. Using different RLs of varying phylogenetic and chemical similarity, we determined that chemosimilarity between donor KRs and recipient KRs correlated with production, in contrast to phylogenetic distance and sequence similarity. Extending our method into multimodular systems that use larger substrates and communication domains, we performed RL swaps in LipPKS2 and found that AP chemosimilarity correlates with production. While our approach did not find a correlation between genetic similarity and production in these diverse RL swaps, it has been shown that within highly similar BGCs, the downstream KS groups with the upstream RL type (e.g., KR, KR-DH, KR-DH- ER).[18] In this study, the donor modules do not share close homology with the lipomycin recipient module, but donor loops with high chemosimilarity located within a BGC may prove more compatible than chemosimilarity alone. Overall, our results determined statistical significance in the correlation between production and the chemosimilarity of the substrate between the donor and recipient modules. More generally, chemoinformatics may provide guideposts for other engineering goals (e.g., KR domain subtype swaps to switch stereo- chemistry). With our incomplete understanding of PKS processing, design principles may accelerate the combinatorial approach currently used for *de novo* biosynthesis and help provide a framework to more rapidly produce valuable biochemicals.

AUTHOR INFORMATION

Corresponding Author

Jay D. Keasling — *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States; QB3 Institute, University of California−Berkeley, Emeryville, California 94608, United States; Department of Chemical &*

Biomolecular Engineering and Department of Bioengineering, University of California, Berkeley, California 94720, United States; Novo Nordisk Foundation Center for Biosustainability, Technical University Denmark, DK2970 Horsholm, Denmark; Synthetic Biochemistry Center, Institute for Synthetic Biology, Shenzhen Institutes for Advanced Technologies, 518055 Shenzhen, China; orcid.org/0000-0003-4170-6088; Email: jdkeasling@lbl.gov

Authors

Amin Zargar − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United*

States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States; QB3 Institute, University of California−Berkeley, Emeryville, California 94608, United States

Ravi Lal − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States*

Luis Valencia − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States*

Jessica Wang − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States*

Tyler William H. Backman − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States*

Pablo Cruz-Morales − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States*

Ankita Kothari − *Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States*

Miranda Werts − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States*

Andrew R. Wong − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States*

Constance B. Bailey − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States; QB3 Institute, University of California−Berkeley, Emeryville, California 94608, United States*

Arthur Loubat − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States*

Yuzhong Liu − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States;* orcid.org/0000-0001-5614-1951

Yan Chen − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States*

Samantha Chang − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States*

Veronica T. Benites − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence*

Berkeley National Laboratory, Berkeley, California 94710, United States; Department of Energy, Agile BioFoundry, Emeryville, California 94608, United States

Amanda C. Hernández − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States*

Jesus F. Barajas − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States; Department of Energy, Agile BioFoundry, Emeryville, California 94608, United States*

Mitchell G. Thompson − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States*

Carolina Barcelos − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States*

Rasha Anayah − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States*

Hector Garcia Martin − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States; Department of Energy, Agile BioFoundry, Emeryville, California 94608, United States; BCAM, Basque Center for Applied Mathematics, 48009 Bilbao, Spain;* orcid.org/0000-0002-4556-9685

Aindrila Mukhopadhyay − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States;* orcid.org/0000-0002-6513-7425

Christopher J. Petzold − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States; Department of Energy, Agile BioFoundry, Emeryville, California 94608, United States;*

orcid.org/0000- 0002-8270-5228

Edward E. K. Baidoo − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, California 94710, United States; Department of Energy, Agile BioFoundry, Emeryville, California 94608, United States*

Leonard Katz − *Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, California 94608, United States; QB3 Institute, University of California−Berkeley, Emeryville, California 94608, United States*

Notes

The authors declare the following competing financial interest(s): J.D.K. has a financial interest in Amyris, Lygos, Demetrix, Napigen, Maple Bio, Berkeley Brewing Sciences, Ansa Biotech, and Apertor Labs.

ACKNOWLEDGMENTS

This work was funded by the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the U.S. Department of Energy (DOE), Office of Science, Office of Biological and Environmental Research, between Lawrence Berkeley National Laboratory and the U.S. Department of Energy, the Agile Biofoundry sponsored by the U.S. DOE Office of Energy Efficiency and Renewable Energy, Bioenergy Technologies and Vehicle Technologies Offices, under Contract DEAC02-05CH11231 between DOE and Lawrence Berkeley National Laboratory, and the National Institutes of Health Awards F32GM125179 and F32GM125166. H.G.M. was also sup- ported by the Basque Government through the BERC 2018- 2021 program and by Spanish Ministry of Economy and Competitiveness MINECO: BCAM Severo Ochoa excellence accreditation SEV-2017- 0718.

REFERENCES

(1) Barajas, J. F.; Blake-Hedges, J. M.; Bailey, C. B.; Curran, S.; Keasling, J. D. Synergy between protein and host level engineering. *Synthetic and Systems Biotechnology* 2017, *2*, 147−166.

(2) Khosla, C.; Herschlag, D.; Cane, D. E.; Walsh, C. T. Assembly line polyketide synthases: mechanistic insights and unsolved problems. *Biochemistry* 2014, *53*, 2875−2883.

(3) Yuzawa, S.; Zargar, A.; Pang, B.; Katz, L.; Keasling, J. D. Commodity chemicals from engineered modular type I polyketide synthases. *Methods Enzymol.* 2018, *608*, 393−415.

(4) Reid, R.; Piagentini, M.; Rodriguez, E.; Ashley, G.; Viswanathan, N.; Carney, J.; Santi, D. V.; Hutchinson, C. R.; McDaniel, R. A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. *Biochemistry* 2003, *42*, 72−79.

(5) Keatinge-Clay, A. Crystal structure of the erythromycin polyketide synthase dehydratase. *J. Mol. Biol.* 2008, *384*, 941−953.

(6) Kellenberger, L.; Galloway, I. S.; Sauter, G.; Böhm, G.; Hanefeld, U.; Cortés, J.; Staunton, J.; Leadlay, P. F. A polylinker approach to reductive loop swaps in modular polyketide synthases. *ChemBioChem* 2008, *9*, 2740−2749.

(7) Hagen, A.; Poust, S.; Rond, T. de; Fortman, J. L.; Katz, L.; Petzold, C. J.; Keasling, J. D. Engineering a polyketide synthase for in vitro production of adipic acid. *ACS Synth. Biol.* 2016, *5*, 21−27.

(8) Gaisser, S.; Kellenberger, L.; Kaja, A. L.; Weston, A. J.; Lill, R. E.; Wirtz, G.; Kendrew, S. G.; Low, L.; Sheridan, R. M.; Wilkinson, B.; Galloway, I. S.; Stutzman-Engwall, K.; McArthur, H. A.; Staunton, J.; Leadlay, P. F. Direct production of ivermectin-like drugs after domain exchange in the avermectin polyketide synthase of Streptomyces avermitilis ATCC31272. *Org. Biomol. Chem.* 2003, *1*, 2840−2847.

(9) Maldonado, A. G.; Doucet, J. P.; Petitjean, M.; Fan, B.-T. Molecular similarity and diversity in chemoinformatics: from theory to applications. *Mol. Diversity* 2006, *10*, 39−79.

(10) Faille, A.; Gavalda, S.; Slama, N.; Lherbet, C.; Maveyraud, L.; Guillet, V.; Laval, F.; Quemard, A.; Mourey, L.; Pedelacq, J.-D. Insights into Substrate Modification by Dehydratases from Type I Polyketide Synthases. *J. Mol. Biol.* 2017, *429*, 1554−1569.

(11) Herbst, D. A.; Jakob, R. P.; Zaffringer, F.; Maier, T. Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. *Nature* 2016, *531*, 533−537.

(12) Barajas, J. F.; McAndrew, R. P.; Thompson, M. G.; Backman, T. W. H.; Pang, B.; de Rond, T.; Pereira, J. H.; Benites, V. T.; Martín, H. G.; Baidoo, E. E. K.; Hillson, N. J.; Adams, P. D.; Keasling, J. D. Structural insights into dehydratase substrate selection for the

borrelidin and fluvirucin polyketide synthases. *J. Ind. Microbiol. Biotechnol.* 2019, *46*, 1225−1235.
(13) McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Betlach, M.; Ashley, G. Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc. Natl. Acad. Sci. U. S. A.* 1999, *96*, 1846−1851.
(14) Chen, X.; Reynolds, C. H. Performance of similarity measures in 2D fragment-based similarity searching: comparison of structural descriptors and similarity coefficients. *J. Chem. Inf. Comput. Sci.* 2002, *42*, 1407−1414.
(15) Peng, H.; Ishida, K.; Sugimoto, Y.; Jenke-Kodama, H.; Hertweck,
C. Emulating evolutionary processes to morph aureothin-type modular polyketide synthases and associated oxygenases. *Nat. Commun.* 2019, *10*, 3918.
(16) Awakawa, T.; Fujioka, T.; Zhang, L.; Hoshino, S.; Hu, Z.; Hashimoto, J.; Kozone, I.; Ikeda, H.; Shin-Ya, K.; Liu, W.; Abe, I. Reprogramming of the antimycin NRPS-PKS assembly lines inspired by gene evolution. *Nat. Commun.* 2018, *9*, 3534.
(17) Eng, C. H.; Backman, T. W. H.; Bailey, C. B.; Magnan, C.; García Martín, H.; Katz, L.; Baldi, P.; Keasling, J. D. ClusterCAD: a computational platform for type I modular polyketide synthase design. *Nucleic Acids Res.* 2018, *46*, D509−D515.
(18) Zhang, L.; Hashimoto, T.; Qin, B.; Hashimoto, J.; Kozone, I.; Kawahara, T.; Okada, M.; Awakawa, T.; Ito, T.; Asakawa, Y.; Ueki, M.; Takahashi, S.; Osada, H.; Wakimoto, T.; Ikeda, H.; Shin-Ya, K.; Abe, I. Characterization of Giant Modular PKSs Provides Insight into Genetic Mechanism for Structural Diversification of Aminopolyol Polyketides. *Angew. Chem., Int. Ed.* 2017, *56*, 1740−1745.
(19) Jenke-Kodama, H.; Börner, T.; Dittmann, E. Natural
biocombinatorics in the polyketide synthase genes of the actino- bacterium Streptomyces avermitilis. *PLoS Comput. Biol.* 2006, *2*, No. e132.
(20) Nguyen, T.; Ishida, K.; Jenke-Kodama, H.; Dittmann, E.; Gurgui, C.; Hochmuth, T.; Taudien, S.; Platzer, M.; Hertweck, C.; Piel. Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat. Biotechnol.* 2008, *26*, 225−233.
(21) Vander Wood, D. A.; Keatinge-Clay, A. T. The modules of trans- acyltransferase assembly lines redefined with a central acyl carrier protein. *Proteins: Struct., Funct., Genet.* 2018, *86*, 664−675.
(22) Yuzawa, S.; Eng, C. H.; Katz, L.; Keasling, J. D. Broad substrate specificity of the loading didomain of the lipomycin polyketide synthase. *Biochemistry* 2013, *52*, 3791−3793.
(23) Tang, Y.; Kim, C.-Y.; Mathews, I. I.; Cane, D. E.; Khosla, C. The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U. S. A.* 2006, *103*, 11124−11129.
(24) Khosla, C.; Tang, Y.; Chen, A. Y.; Schnarr, N. A.; Cane, D. E. Structure and mechanism of the 6-deoxyerythronolide B synthase. *Annu. Rev. Biochem.* 2007, *76*, 195−221.
(25) Yan, J.; Hazzard, C.; Bonnett, S. A.; Reynolds, K. A. Functional modular dissection of DEBS1-TE changes triketide lactone ratios and provides insight into Acyl group loading, hydrolysis, and ACP transfer. *Biochemistry* 2012, *51*, 9333−9341.
(26) Zotter, A.; Baüerle, F.; Dey, D.; Kiss, V.; Schreiber, G. Quantifying enzyme activity in living cells. *J. Biol. Chem.* 2017, *292*, 15838−15848.
(27) Phelan, R. M.; Sachs, D.; Petkiewicz, S. J.; Barajas, J. F.; Blake- Hedges, J. M.; Thompson, M. G.; Reider Apel, A.; Rasor, B. J.; Katz, L.; Keasling. Development of Next Generation Synthetic Biology Tools for Use in Streptomyces venezuelae. *ACS Synth. Biol.* 2017, *6*, 159−166.
(28) Weber, A. L. Origin of fatty acid synthesis: thermodynamics and kinetics of reaction pathways. *J. Mol. Evol.* 1991, *32*, 93−100.